The Thesis Committee for Ricky Wayne Aldon Smith
Certifies that this is the approved version of the following thesis:

Assessing the preservation of cytosine methylation in ancient DNA from five prehistoric Native American populations

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Deborah A. Bolnick

Anthony DiFiore

Assessing the preservation of cytosine methylation in ancient DNA from five prehistoric Native American populations

by

Ricky Wayne Aldon Smith, B.S.; M.S.

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

May 2013

Acknowledgements

I thank Cara Monroe, Fernando Villanea, Brian Kemp, and Jaime Mata-Miguez for contributing skeletal remains or aDNA extracts from the Yukisma and Xaltocan sites for this study. I also thank my committee members, Deborah Bolnick and Anthony DiFiore, for their invaluable contributions to this work.

Abstract

Assessing the preservation of cytosine methylation in ancient DNA from five prehistoric Native American populations

Ricky Wayne Aldon Smith, M.A.

The University of Texas at Austin, 2013

Supervisor: Deborah A. Bolnick

Cytosine methylation of CpG dinucleotides is an important epigenetic mark that regulates gene expression in humans. While methylation patterns in extant populations have been widely studied, few studies have attempted to analyze methylation in ancient DNA. Indeed, it was only recently shown that methyl groups can be preserved in ancient DNA. However, it is unknown how often methylation patterns can be recovered from ancient samples with preserved nuclear DNA. If they are frequently preserved, it may ultimately be possible to infer patterns of gene activity at the population level in ancient times.

In this study, I assessed the preservation of cytosine methylation in ancient DNA from the remains of 30 prehistoric Native Americans from California, Illinois, Kentucky, and Mexico. These samples were previously shown to contain endogenous mitochondrial and nuclear DNA. I analyzed the cytosine methylation states of CpG-rich retrotransposons, which are epigenetically inactivated by cytosine methylation in humans. Unmethylated cytosines were converted to uracils by treatment with sodium

bisulfite. Bisulfite products were pyrosequenced, and C-to-T conversions at potentially methylated CpG dinucleotides were quantified. I found that cytosine methylation is readily recoverable from human remains with preserved nuclear DNA from various localities over the time depth tested (~6000 years). This study presents the first direct evidence of cytosine methylation in ancient human remains, and suggests that it may be possible to analyze patterns of gene activity in ancient populations.

Table of Contents

List of Tables

List of Figures

Introduction

Epigenetic marks comprise a variety of stable, chemical modifications to DNA and its associated proteins that influence chromatin structure and regulate gene expression. These marks designate which genomic segments are available for transcription, providing a means for regulating gene activity without changing the underlying nucleotide sequence (Attwood et al. 2002). Some epigenetic patterns, and their associated spectrums of gene expression, can be maintained across cell divisions and even across generations, allowing certain patterns of gene regulation to be inherited. Functionally, epigenetic gene regulation plays a crucial role in development and in mediating gene-by-environment interactions, and underlies complex diseases (Reik 2007, Jaenisch and Bird 2003, Portela and Estellar 2010).

One widely studied type of epigenetic mark is cytosine methylation, a DNA-level modification in which a methyl group is added to a cytosine. In humans and other mammals, methylation occurs most often at cytosines found in CpG dinucleotide contexts, where a cytosine and guanine occur contiguously along the DNA backbone. When CpGs are methylated, the methyl group protrudes into the major groove of the DNA, often preventing active transcription and promoting chromatin formation. Changes in cytosine methylation patterns have consequences for DNA transcription, providing a plastic, pre-transcriptional means of modifying gene expression. The positions of CpG dinucleotides in the genome, therefore, are functionally relevant to gene expression.

Methylated cytosines (5-methylcytosine or 5mC) in CpG dinucleotide contexts are vastly underrepresented in the human genome compared to other nucleotide bases and dinucleotide combinations (Sved and Bird 1990), and are often concentrated in regions of

high density such as CpG islands and repetitive elements. Islands are highly enriched in CpG dinucleotides compared to the rest of the genome, and often occur near gene promoters and transcriptional start sites, where they serve as genomic landmarks and epigenetic control sites for transcriptional activation or inactivation (Gardiner-Garden and Frommer 1987, Takai and Jones 2002). When the methylation status of a CpG island changes, it can affect the transcription of associated genes. CpG islands lying in or near the transcriptional start sites of genes are most commonly unmethylated, while CpGs found outside promoters are most often methylated. Other relatively CpG-rich regions of the genome include retrotransposable elements such as Long Interspersed Elements (LINEs) and Short Interspersed Elements (SINEs), which are usually epigenetically inactivated through cytosine methylation to prevent aberrant transposition (Gama-Sosa 1983, Bestor 1998).

While the effects of cytosine methylation on gene expression have been widely studied in extant species, few studies have attempted to analyze epigenetic marks in the DNA of ancient or extinct organisms. However, analyzing the epigenetic regulation of DNA within archaeological, paleoanthropological, or paleontological contexts could reveal patterns of gene expression in the past, making it possible to reconstruct some types of gene activity in extinct species with preserved DNA. Gene expression studies could therefore become feasible for archaic hominins such as Neanderthals and Denisovans, as well as fauna like mammoth, bison, muskox, cave bear, and moa, which have all yielded endogenous nuclear aDNA (Green et al 2010, Reich et al 2010, Miller et al 2008, Debruyne et al. 2008, Llamas et al 2012, Campos et al. 2010, Calvignac et al. 2008, Stiller et al. 2010, Orlando et al. 2002, Hofreiter et al. 2002, McCallum et al 2013).

However, post-mortem damage to aDNA, including hydrolytic lesions, has been a major concern for sequence accuracy when working with ancient sources of DNA

(Hofreiter et al 2001, Briggs et al 2010). Certain classes of cytosine damage may present additional challenges to the accurate measurement of cytosine methylation in ancient DNA (aDNA). When unmethylated cytosines are deaminated after an organism's death, they degrade to uracils, while deaminated 5mCs degrade to thymines. Both nucleotide base changes result in the misincorporations of thymines during polymerase chain reaction (PCR) amplification and sequencing, and tracing the original methylation status of cytosines could be difficult.

Recent studies indicate that cytosine methylation is preserved in some ancient specimens. Briggs and colleagues (2010) found indirect evidence for cytosine methylation in aDNA extracted from 43,000-year-old Mammuthus primigenius and 38,000-year-old Homo sapiens neanderthalensis remains. In this study, a uracil-DNA-glycosylase (UDG) and endonuclease VIII (endoVIII) protocol was used to repair aDNA extracts prior to sequencing, in an effort to reduce sequencing errors resulting from post-mortem damage to aDNA. This treatment removes uracils formed by the deamination of unmethylated cytosines, greatly increasing sequence accuracy in aDNA sequencing by reducing C/G→T/A conversions. However, the researchers observed an incomplete rescue of C/G→T/A misincorporations using their repair protocol, and found that unrepaired base misincorporations were concentrated in CpG contexts of nuclear DNA. They suggested that these misincorporations may be linked to the methylation of cytosines in CpG contexts because the deamination of 5mC results in thymine, a base misincorporation error that is not repaired by the UDG-endoVIII approach.

More recently, cytosine methylation has been directly detected in aDNA recovered from the remains of a single, 26,000-year-old Bison priscus specimen (Llamas et al. 2012). This study analyzed four repetitive genomic elements and two single-copy imprinted genes, and used bisulfite allelic sequencing to detect cytosine methylation.

Bisulfite sequencing is a commonly used method for analyzing cytosine methylation that relies on the conversion of unmethylated cytosines to uracils (which are sequenced as thymines), where the remaining cytosines detected during sequencing are methylated cytosines resistant to conversion by sodium bisulfite. Llamas et al. (2012) compared methylation patterns from the extinct specimen with homologues in contemporary Bos taurus to establish the validity of the methylation patterns observed in the aDNA sequences. However, while Llamas and colleagues (2012) showed that methylation patterns can be preserved in aDNA, they reported methylation results for only one of the six ancient specimens tested. Therefore, it is unknown how often cytosine methylation can be recovered from ancient specimens with preserved nuclear DNA. Whether this solitary result reported by Llamas and colleagues is unique, or whether methylation may be frequently preserved in ancient samples with preserved nuclear DNA, has yet to be evaluated.

In this study, I evaluated the preservation of cytosine methylation in 30 prehistoric human remains from five geographical localities, ranging in age from 200 to 6000 years before present (ybp). To assess the degree of cytosine methylation preservation in these samples, human-specific LINE-1 (L1Hs) elements were targeted. These CpG-rich repetitive elements are known to be epigenetically inactivated by cytosine methylation as a host-defense mechanism against aberrant transposition. While expression of LINE-1 elements in somatic tissues has occasionally been observed (Munoz-Lopez et al. 2012), particularly during development, their active transcription is generally regarded as atypical because random retrotransposition of LINEs into gene sequences can have deleterious effects. When LINE promoters are methylated, mRNA expression is significantly reduced, especially when that methylation is concentrated in the 5' promoter sequence (Yoder JA, Walsh CP, Bestor 1997). Thus, because LINE-1

elements are epigenetically inactivated by cytosine methylation, they make ideal targets for the analysis of cytosine methylation in aDNA.

The LINE insertion examined in this study, L1Hs56 (GenBank: AC005908), is a member of the transcribed, subset a (Ta) family of L1Hs retrotransposons. The Ta family of L1 elements was selected for this study because it is the most recently integrated subfamily of L1 elements in the human genome (Kazazian and Moran 1998; Boissinot et al. 2000; Sheen et al. 2000, Myers et al. 2002). As 5mCs are highly mutable, they are gradually depleted from CpG rich regions of the genome over evolutionary time. Thus, L1Hs families that were integrated in the more distant past may not be CpG dense enough to provide ideal targets for methylation analysis. Conversely, the more recently integrated families, while retaining sufficient CpG density, are often polymorphic due to incomplete lineage sorting and may not be found in all humans. Therefore, fixed loci within the Ta subfamily represent ideal targets for methylation analysis because they were integrated recently enough to retain sufficient CpG density and are monomorphic among humans. Taken together, these characteristics make L1Hs (Ta) elements excellent targets for a diagnostic designed to detect the preservation of methylation patterns in aDNA. L1Hs56, the LINE analyzed in this study, is monomorphic in human populations and consistently methylated in normal, somatic tissues, allowing us to assess how often 5mCs are preserved at this locus in ancient human skeletal remains (Myers et al. 2002).

Here, I report the direct detection of cytosine methylation in prehistoric Native American remains from five geographical localities across the Americas, spanning a time interval of more than 6000 years. The aDNA samples selected for this study previously amplified both nuclear and mitochondrial loci (Bolnick and Smith 2007; Mata-Míguez et al. 2012; Villanea et al., in press; Bolnick and Bonine, unpublished data), demonstrating that both sources of DNA are well preserved. This study presents the first direct evidence

of cytosine methylation in human aDNA and suggests that it may be possible to analyze some patterns of gene activity in ancient populations.

Materials and Methods

DNA Samples

The aDNA extracts used in this study were previously obtained from Native American skeletal remains, and previous analyses demonstrated the preservation of endogenous aDNA in these samples through the successful amplification of both mitochondrial and nuclear loci (Bolnick and Smith 2007; Mata-Míguez et al. 2012; Villanea et al., in press; Bolnick and Bonine, unpublished data). In total, 30 samples were drawn from five distinct localities throughout North and Central America (Table 1), and range in age from approximately 200 to 6000 years before present (ybp).

To contextualize the aDNA methylation data and to identify any differences in methylation statuses among the ancient samples that might be due to post-mortem DNA degradation, DNA was extracted from seven living individuals for comparison (Table 2). Both ancient and contemporary sources of DNA were analyzed using the same protocol. The extraction and analysis of DNA in this study was approved by the University of Texas at Austin's Institutional Review Board (protocol #2012-05-0105).

Table 1: Provenience and specifications for human remains used in this study

Sample ID	**Extraction method**	**Locality**	**Cultural Affiliation**	**Site Date**	**Reference**
RIC6	Phenol-Chloroform	Rickets Mound, Kentucky	Adena	2500-1800 BP	Bolnick and Bonine, unpublished data
RIC10	Phenol-Chloroform	Rickets Mound, Kentucky	Adena	2500-1800 BP	Bolnick and Bonine, unpublished data
RIC13	Phenol-Chloroform	Rickets Mound, Kentucky	Adena	2500-1800 BP	Bolnick and Bonine, unpublished data
RIC14	Phenol-Chloroform	Rickets Mound, Kentucky	Adena	2500-1800 BP	Bolnick and Bonine, unpublished data
RIC15	Phenol-Chloroform	Rickets Mound, Kentucky	Adena	2500-1800 BP	Bolnick and Bonine, unpublished data
WSU23	Silica, GITC	Yukisma Site, California	Muwekma Ohlone	740-230 BP	Villanea et al., in press
WSU43	Silica, GITC	Yukisma Site, California	Muwekma Ohlone	740-230 BP	Villanea et al., in press
WSU50	Silica, GITC	Yukisma Site, California	Muwekma Ohlone	740-230 BP	Villanea et al., in press
WSU134	Silica, GITC	Yukisma Site, California	Muwekma Ohlone	740-230 BP	Villanea et al., in press
WSU170	Silica, GITC	Yukisma Site, California	Muwekma Ohlone	740-230 BP	Villanea et al., in press
E102	Silica, GITC	Xaltocan, Mexico	Otomi/Aztec	700-400 BP	Mata-Miguez et al. 2012
E304	Silica, GITC	Xaltocan, Mexico	Otomi/Aztec	700-400 BP	Mata-Miguez et al. 2012
E512	Silica, GITC	Xaltocan, Mexico	Otomi/Aztec	700-400 BP	Mata-Miguez et al. 2012
E85	Silica, GITC	Xaltocan, Mexico	Otomi/Aztec	700-400 BP	Mata-Miguez et al. 2012
Y222	Silica, GITC	Xaltocan, Mexico	Otomi/Aztec	700-400 BP	Mata-Miguez et al. 2012
Y23	Silica, GITC	Xaltocan, Mexico	Otomi/Aztec	700-400 BP	Mata-Miguez et al. 2012
Y252	Silica, GITC	Xaltocan, Mexico	Otomi/Aztec	700-400 BP	Mata-Miguez et al. 2012
Y27	Silica, GITC	Xaltocan, Mexico	Otomi/Aztec	700-400 BP	Mata-Miguez et al. 2012
Y341	Silica, GITC	Xaltocan, Mexico	Otomi/Aztec	700-400 BP	Mata-Miguez et al. 2012
Y391	Silica, GITC	Xaltocan, Mexico	Otomi/Aztec	700-400 BP	Mata-Miguez et al. 2012
K51M	Phenol-Chloroform	Pete Klunk Mound Group, Illinois	Middle Woodland (Hopewell)	1825 BP	Bolnick and Smith 2007
K515	Phenol-Chloroform	Pete Klunk Mound Group, Illinois	Middle Woodland (Hopewell)	1825 BP	Bolnick and Smith 2007
K524	Phenol-Chloroform	Pete Klunk Mound Group, Illinois	Middle Woodland (Hopewell)	1825 BP	Bolnick and Smith 2007
K529	Phenol-Chloroform	Pete Klunk Mound Group, Illinois	Middle Woodland (Hopewell)	1825 BP	Bolnick and Smith 2007
K1159	Phenol-Chloroform	Pete Klunk Mound Group, Illinois	Middle Woodland (Hopewell)	1825 BP	Bolnick and Smith 2007
K618	Phenol-Chloroform	Pete Klunk Mound Group, Illinois	Middle Woodland (Hopewell)	1825 BP	Bolnick and Smith 2007
OH-93	Silica, GITC	Indian Knoll, Kentucky	Green River	6000-4500 BP	Bolnick and Bonine, unpublished data
OH-98	Silica, GITC	Indian Knoll, Kentucky	Green River	6000-4500 BP	Bolnick and Bonine, unpublished data
OH-266	Silica, GITC	Indian Knoll, Kentucky	Green River	6000-4500 BP	Bolnick and Bonine, unpublished data
OH-298	Silica, GITC	Indian Knoll, Kentucky	Green River	6000-4500 BP	Bolnick and Bonine, unpublished data

GITC – Guanidinium thiocyanate.

Table 2: Percent Methylation of L1Hs56 for Contemporary Human Samples

Sample	% Methylation		
	PCR 1	PCR 2	PCR3
1	61	60	60
2	61	58	62
3	61	60	64
4	50	54	50
5	56	55	56
6	52	50	52
7	57	60	58

BISULFITE CONVERSION

Aliquots of the aDNA extracts were prepared for methylation analysis using the EpiTect Bisulfite Kit (Qiagen), following the manufacturer's protocol for converting unmethylated cytosines in small amounts of fragmented DNA. Five µL of aDNA from each extract were used in a 140 µL bisulfite conversion reaction. This protocol converts unmethylated cytosines to uracils by treatment with sodium bisulfite (NaHSO3). Uracils are subsequently incorporated as thymines during target amplification, so any cytosines detected during sequencing represent methylated cytosines, which are resistant to sodium bisulfite conversion.

DNA AMPLIFICATION, SEQUENCING, AND ANALYSIS

Primers for PCR amplification and pyrosequencing of L1Hs56 in bisulfite converted DNA were designed using the Qiagen Q24 PyroMark Assay Design Software, version 2.0. The primers amplify an 87 base pair (bp) fragment of L1Hs56 (forward primer: 5'-AGTAAAGTTTTTAAGAAATATGGGATTATG, biotinylated reverse primer: 5'-biotin- TTCCATTCTCCACATCACTTTCAAATAC). PCRs were prepared with 2 µL of bisulfite product in a 15 µL reaction volume using Qiagen's PyroMark PCR kit. ? µL of 20 mg/mL BSA (Roche) and ? µL of MasterAmpTM 10X PCR enhancer with betaine (Epicentre) were also included in the reaction. PCR conditions included an initial denaturation at 95°C for 15 minutes, 58 cycles of denaturation at 94°C for 30

seconds, annealing at 58°C for 30 seconds, and extension at 72°C for 30 seconds, and a final extension at 72°C for 10 minutes.

Following amplification, 3 μL of the resulting amplicons were visualized with GelRed on a 6% polyacrylamide gel to confirm amplification prior to pyrosequencing. The remaining PCR product of each confirmed amplicon was submitted to the DNA Sequencing Facility at the University of Texas at Austin for pyrosequencing and CpG analysis on the Qiagen Q24 platform. Pyrosequencing reactions were initiated with a sequencing primer (5'ATATGGGATTATGTGAAAAG) that targets an internal fragment of the 87 bp amplicon and generates a 16 bp read containing a single CpG site. Comparison of C→T transitions at the target CpG position was used to calculate percent methylation values for each sample, where Cs represent the proportion of methylated cytosines and Ts represent the proportion of unmethylated cytosines following bisulfite conversion.

EVALUATION OF ANCIENT DNA QUALITY

To evaluate whether overall aDNA quality (e.g., DNA concentration and range of fragment sizes in the original DNA extracts) was correlated with the percent methylation observed for each sample, I submitted a subset of aDNA extracts for fragment size analysis and quantitation on the Agilent 2100 Bioanalyzer platform at the University of Texas at Austin DNA Sequencing Facility. The 11 samples selected for this analysis had percent methylation data from at least two independent pyrosequencing runs. Six of these samples exhibited low variability in percent methylation between runs, while five exhibited high variability in percent methylation between runs.

STATISTICAL ANALYSES

To evaluate whether any of the observed variation in percent methylation values was statistically significant, analyses of variance (ANOVAs) were performed using the [R] statistical platform (R Core Team 2013). Three sets of analyses were conducted. First, samples were grouped according to locality and percent methylation values were compared to assess the influence of geographic location on methylation signal. Second, all ancient samples were grouped together and compared against all contemporary samples to evaluate whether the range of percent methylation values in ancient degraded samples fell within the expected range, with the contemporary data being used to define expectations for cytosine methylation levels in non-degraded DNA.

CONTAMINATION CONTROLS

The analysis of highly degraded DNA from prehistoric human remains requires strict precautions to minimize contamination from exogenous sources of DNA and to detect any contamination that may occur (Kaestle and Horsburgh 2002, Paabo 2004, Gilbert et al. 2005, Willerslev and Cooper 2005). To this end, bisulfite conversion of aDNA extracts and PCR setups were performed in the aDNA facility at the University of Texas at Austin. The aDNA facility is a restricted-access, positive air pressure, HEPA-filter ventilated space with overhead UV-irradiating lights that is dedicated to pre-PCR analyses of aDNA. The post-PCR facility is located in a separate building, and all movement of materials and personnel was unidirectional (from pre-PCR to post-PCR facility) to prevent contamination from DNA amplicons. Additional precautions included the use of sterile and disposable hooded coveralls, hair covers, face masks, sleeve covers, dedicated shoes, and two pairs of gloves. Laboratory benchtops and equipment were frequently decontaminated with 6% sodium hypochlorite (full strength bleach), and the entire lab was decontaminated weekly with a 3% sodium hypochlorite solution

(1:1::bleach:water, v/v). The facility was also irradiated with 254-nm emitting UV light for 12 hours following each use, while tubes, containers, and reagents (when possible) were UV irradiated in a 254-nm emitting DNA cross-linker for 15 minutes prior to use. Negative controls included conversion blanks (bisulfite conversion reaction mixtures containing no DNA) and PCR amplification negatives to identify the presence of any contamination at each stage of sample analysis. Finally, two independent PCR amplifications were performed for each DNA extract, and where sufficient material was available, additional PCR amplifications of an independent DNA extraction were performed to verify the authenticity of the results.

Results

Of the 30 aDNA samples examined in this study, all samples produced at least one successful amplification of L1Hs56 from which percent methylation could be determined (Tables 1 and 2; Figure 1A). However, in three samples where second extracts were available (RIC6, K5-1, and K618), the second extract never yielded amplification of L1Hs56. Furthermore, for four samples with only a single extract (K5-24, K5-29, OH93, Y391), only a single amplicon could be generated for which percent methylation could be determined. In total, 23 samples yielded reproducible amplicons from the same extract, and of these, 10 successfully amplified at least once from a second extract. Overall, percent methylation at the variable CpG position in the 87 base pair L1Hs56 amplicon ranged between 37% and 77% in the ancient samples, with an average value of 52% methylation. For the # contemporary samples, all samples successfully amplified for all three replicate PCRs, and percent methylation ranged between 50% and 64%, with an average value of 57% (Table 3).

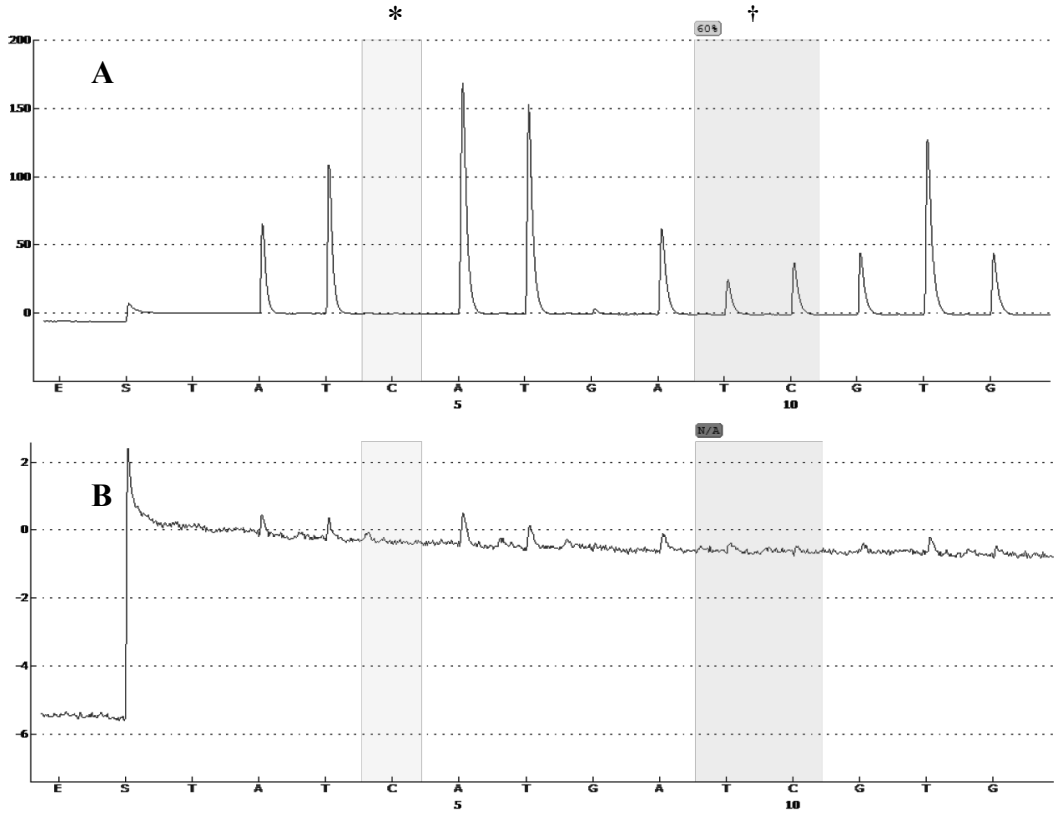
Table 3: Percent Methylation of L1Hs56 for Prehistoric Human Samples

Sample	Locality	Date	Extract	% Methylation	
				PCR 1	PCR 2
RIC6	Ricketts Mound, Kentucky	2500-1800 BP	1	69	60
			2	NA	NA
RIC10	Ricketts Mound, Kentucky	2500-1800 BP	1	80	NA
			2	55	52
RIC13	Ricketts Mound, Kentucky	2500-1800 BP	1	51	50
RIC14	Ricketts Mound, Kentucky	2500-1800 BP	1	47	41
			2	50	52
RIC15	Ricketts Mound, Kentucky	2500-1800 BP	1	50	51
			2	44	50
WSU23	Yukisma Site, California	740-230 BP	1	49	50
WSU43	Yukisma Site, California	740-230 BP	1	50	49
WSU50	Yukisma Site, California	740-230 BP	1	50	47
WSU134	Yukisma Site, California	740-230 BP	1	53	50
WSU170	Yukisma Site, California	740-230 BP	1	53	52
K5-1	Pete Klunk Mound, Illinois	1825 BP	1	50	NA
			2	NA	NA
K5-15	Pete Klunk Mound, Illinois	1825 BP	1	50	50
			2	60	NA
K5-24	Pete Klunk Mound, Illinois	1825 BP	1	70	NA
K5-29	Pete Klunk Mound, Illinois	1825 BP	1	48	NA
K11-59	Pete Klunk Mound, Illinois	1825 BP	1	47	45
K618	Pete Klunk Mound, Illinois	1825 BP	1	77	NA
			2	NA	NA
OH93	Indian Knoll, Kentucky	6000-4500 BP	1	65	NA
OH98	Indian Knoll, Kentucky	6000-4500 BP	1	50	58
OH266	Indian Knoll, Kentucky	6000-4500 BP	1	37	48
OH298	Indian Knoll, Kentucky	6000-4500 BP	1	55	58
E102	Xaltocan, Mexico	700-400 BP	1	52	54
			2	51	55
E304	Xaltocan, Mexico	700-400 BP	1	46	NA
			2	42	53
E512	Xaltocan, Mexico	700-400 BP	1	47	NA
			2	48	60
E85	Xaltocan, Mexico	700-400 BP	1	59	63
			2	56	55
Y22	Xaltocan, Mexico	700-400 BP	1	46	45
Y23	Xaltocan, Mexico	700-400 BP	1	65	66
			2	55	44
Y252	Xaltocan, Mexico	700-400 BP	1	48	40
Y27	Xaltocan, Mexico	700-400 BP	1	51	60
			2	53	51
Y341	Xaltocan, Mexico	700-400 BP	1	43	43
Y391	Xaltocan, Mexico	700-400 BP	1	70	NA

NA – No amplification.

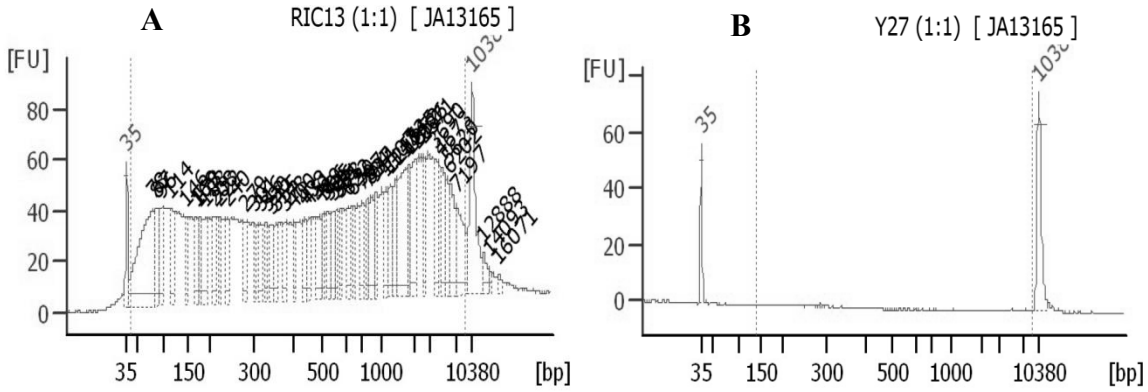
Bisulfite control positions in each pyrogram showed no cytosine misincorporations (Figure 1), indicating the efficient conversion of unmethylated cytosines to uracils for both ancient and contemporary sources of DNA used in this study. Control dispensations (dispensation 7) indicated negligible levels of background noise in the pyrosequencing reaction (Figure 1). Finally, both the bisulfite conversion blanks and PCR amplification blanks showed no amplification, demonstrating that no exogenous DNA contamination was introduced during bisulfite conversion or PCR setup (Figure 1B).

Figure 1: Representative pyrograms for (*A*) an aDNA sample (Ricketts, burial 6) and (*B*) no-template control. *Bisulfite control position. †Methylated CpG target position.



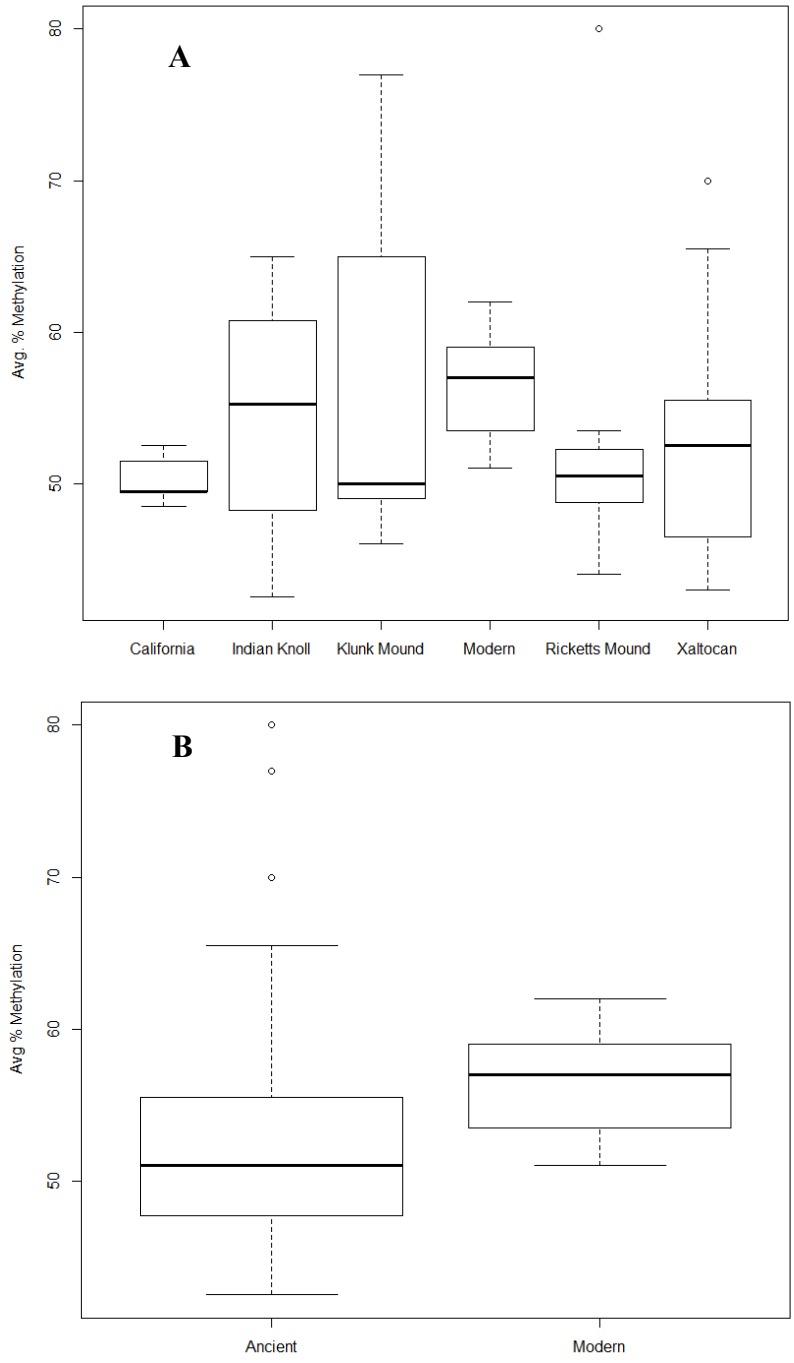
To help evaluate how aDNA preservation might be related to the percent methylation values generated for these samples, I estimated the DNA concentration and range of fragment sizes in 11 aDNA extracts. Generally, samples with low variance in percent methylation values between pyrosequencing runs have higher concentrations of DNA across a wider range of short fragment lengths than samples with high variance between runs, which have little detectable aDNA (Figure 2). While these results are not conclusive, owing to the expected and confounding presence of exogenous contaminants such as co-extracted bacterial genomes, the BioAnalyzer results do indicate a possible relationship between aDNA quality and percent methylation patterns.

Figure 2: BioAnalyzer electropherograms for representative samples with (*A*) low and (*B*) high variance in percent methylation values across pyrosequencing runs.



Finally, one-way ANOVAs accounting for repeated measures were performed to test for statistical differences in percent methylation between localities. Additional ANOVAs compared ancient and contemporary cohorts to assess whether the methylation signals in ancient samples fell within the expected range, as established by contemporary data. In comparing percent methylation data across the various localities, marked differences in the means and variances are apparent. For example, the Indian Knoll, Xaltocan, and Klunk Mound groups show wide variations in percent methylation and notable differences in mean percent methylation, while the California and Ricketts Mound groups show comparatively narrower variation in percent methylation. However, because the ranges of variation overlap across sites, the ANOVA results indicate no statistically significant differences ($P = 0.75$, $df=4$) in percent methylation data between the five localities studied (Figure 3A). When all the samples were divided into two temporal groupings (ancient and contemporary), a decrease in the average percent methylation was observed (Figure 3B). However, this difference is not statistically significant ($P= 0.42$, $df=1$).

Figure 3: ANOVAs comparing % methylation (*A*) by site, and (*B*) between ancient and contemporary cohorts.



Discussion

This study presents the first direct evidence that cytosine methylation is preserved in aDNA extracted from prehistoric human remains. L1Hs56 was successfully amplified and percent methylation was determined in samples from five distinct geographical localities spanning a time period of 6000 years. This study builds on previous efforts to detect 5mC in aDNA, which hinged on a solitary observation of cytosine methylation in Bison priscus. The data reported here demonstrate that cytosine methylation is readily recoverable from prehistoric samples with preserved nuclear DNA. The results in this study are thought to be derived from endogenous ancient DNA because of strict adherence to laboratory precautions and controls, multiple independent verifications of percent methylation data for each sample, and an absence of contaminant DNA in all control reactions. Given the high success rate in this study, it is clear that ancient samples with preserved nuclear DNA, though already heavily fragmented by post-mortem processes, can maintain sufficient concentrations of DNA for analysis following the harsh effects of exposure to sodium bisulfite, which can reduce the number of viable molecules by ~90% or more (Grunau et al 2001).

The choice of a locus with a known methylation status is critical when initially evaluating whether cytosine methylation is preserved in aDNA. Methylation within single copy loci such as differentially methylated regions (DMRs) and promoters can show variable levels of methylation, owing to functional differences in epigenetic gene regulation. Within an aDNA context, it would be more difficult to determine whether the observed variation in these single copy loci is due to normal variation in gene expression, or due to cytosine damage stemming from post-mortem degradation. L1Hs56 was chosen for this study because it is known to be consistently methylated in healthy,

somatic tissues (Gama-Sosa 1983, Bestor 1998). The choice of a locus of known methylation status helps ensure that the observed variation in methylation levels between samples is not the result of differences in individual levels of gene expression. Instead, L1Hs56 serves as a control for differences in gene expression because it is consistently epigenetically inactivated by cytosine methylation. Thus, the inability to detect cytosine methylation, or variation in detectable levels of methylation between samples, can be more directly associated with the issue of DNA preservation itself.

Sample preservation can present additional challenges to measuring cytosine methylation in aDNA in terms of the quality of template material. Indeed, an association between overall DNA preservation and percent methylation values was observed. The BioAnalyzer data indicated that samples with low variability in percent methylation between pyrosequencing runs have higher concentrations of DNA fragments over a wider range of fragment lengths, while samples with high variability between runs have markedly lower concentrations of DNA. Thus, samples with better DNA preservation yield more consistent signals of percent methylation between independent pyrosequencing runs. This observation is likely due to the fact that the percent methylation values obtained through pyrosequencing are averaging over a greater number of template molecules in well preserved samples, while percent methylation values are prone to wider swings in poorly preserved samples because of bias towards a relatively smaller number of starting templates. For samples from Xaltocan, Mexico, the relationship between sample preservation and variance in percent methylation values is further supported by data on 15 autosomal short tandem repeat (STR) loci, where samples with lower variances (better DNA preservation) have been successfully used to amplify longer STR fragments (up to 450 bp) than samples with higher variances (Mata-Míguez and Bolnick, unpublished data).

Variation in percent methylation values is apparent in the ANOVAs comparing the five prehistoric localities, and when comparing all ancient samples against all contemporary samples. However, no statistically significant differences were observed in percent methylation across the various localities. Because each locality represents a different time period (ranging from 200 to 6000 ybp) and because each set of samples will be associated with different preservational contexts, the "locality" designation is a conflation of the various potential effects these different factors have on percent methylation. Likewise, when all ancient samples were compared against all contemporary samples, the average percent methylation is lower and the variance more widely distributed than in contemporary samples, but again, this difference is not statistically significant. While these data do not statistically distinguish between groups, it should be noted that the sample sizes are relatively small, and there may be a lack of statistical power to distinguish potential differences. The observed differences may therefore still represent biologically meaningful differences.

There are at least three reasons why different averages and wide variances might be observed in the percent methylation data from ancient specimens. First, the decrease in percent methylation among ancient specimens might be due to post-mortem depurination of methylated cytosines to thymines. Because degraded 5mCs are indistinguishable from unmethylated cytosines following bisulfite conversion and PCR amplification, deamination of 5mCs may artificially decrease the methylation signal from its original status in living organisms. If, however, damage can be ruled out as a potential confounding factor, a second possible explanation is the relationship between nutritional status and genome-wide levels of cytosine methylation. It has been recently shown that a lack of methyl donors in the diet can lead to global genomic hypomethylation in a population, and this effect is independent of any specific pathology (Zhang et al 2011).

Thus, the observed differences in percent methylation between ancient and contemporary samples may be genuine, and result from underlying differences in the nutritional status of ancient and contemporary populations. Finally, it is possible that the differences in percent methylation values seen here represent normal variation for LINE1 insertions, and lack any particular biological significance.

In the present study, it seems unlikely that cytosine damage has played a major role in decreasing the percent methylation detected among the ancient samples. Depurination has been associated with points of strand breakage in aDNA, and as a result, deaminated cytosines are highly concentrated at 5' ends (Briggs et al 2007). Since my target CpG lies in the center of the 87 bp portion of the L1Hs56 sequence, PCR amplicons would be generated from only the proportion of fragments that are intact over the nucleotide sequences immediately surrounding the cytosine of interest. Therefore, given the features of the current protocol, it is unlikely that any decreases to percent methylation values are due to post-mortem processes. The observed differences in percent methylation between sites, and between ancient and contemporary cohorts, then, may represent variability in nutritional status or naturally occurring variation in L1Hs methylation patterns. However, the current study was not designed to distinguish the underlying causes of variation in percent methylation, and more research will be needed to distinguish among these possible explanations.

Conclusion

This study presents the first direct detection of cytosine methylation in ancient human remains. Cytosine methylation is readily recoverable from prehistoric skeletal material with preserved mitochondrial and nuclear DNA from five distinct localities spanning a time depth of 6000 years. These results provide the first evidence that methylation is preserved well enough in many aDNA samples that it may be possible to study epigenetic gene regulation and infer gene expression levels in populations of ancient or extinct organisms. However, the potential for epigenome reconstruction will be limited by the availability of prehistoric source materials. Soft tissues are rarely available, and where they are available, DNA preservation is often poor. Thus, epigenome analysis will largely be limited to teeth and bone, which will contain epigenomes relevant only to bone and peripheral blood tissues.

Anthropological Significance

The finding that cytosine methylation is commonly preserved in aDNA opens up the potential for innovative analyses in anthropological genetics, a field that has conventionally been concerned with sequence-based analyses aimed at reconstructing population histories, reconstructing phylogenies, and inferring rates of molecular evolution (Kaestle and Horsburgh 2002). The analysis of methylation patterns in aDNA may eventually be used to infer patterns of gene expression in the past, which can be used to reconstruct how patterns of gene regulation have changed over time. Such studies could shed light on biological phenomena that cannot be detected through the analysis of DNA sequences, such as (1) changes in epigenetic gene regulation that drive local adaptation, and (2) inferring the effects of social and ecological environments from patterns of gene expression. Both of these examples will be considered here.

Epigenetic marks are known to vary in response to certain kinds of environmental stimuli, and thus, they play a central role in mediating gene-by-environment interactions and in allowing variable gene expression in different environments. In some cases, these environmentally-induced changes in epigenetic gene expression can be inherited from generation to generation. The observation that some methylation patterns show evidence of intergenerational inheritance raises the potential for natural selection to act on variation in patterns of gene regulation to drive local adaptations over shorter time scales than are possible from nucleotide substitutions within regulatory elements (Guerrero-Bosagna et al. 2005). Epigenetic changes giving rise to local adaptations raise many interesting possibilities for novel research within anthropological genetics, including improving our understanding of the process of divergence between closely related hominids like Neanderthals, Denisovans, and modern humans.

These closely related species share a high sequence identity due to their relatively recent divergence from a common ancestor. Thus, at least some of the important differences driving inter-species differences may be regulatory in nature, rather than driven by more widespread sequence diversification (Noonan 2010). While previous research has yielded interesting speculations about the evolutionary differences between humans and archaic hominins based on functional gene variants (Green et al 2010), the ability to measure epigenetic gene regulation in aDNA of these extinct hominids would make it possible to study other important inter-species differences that may not be evident from sequence or morphological measures of divergence, such as epigenetic changes to gene expression that affect development.

Finally, studies of gene-by-environment changes in epigenetic regulation could be used to investigate questions about the social and ecological experiences of ancient humans. For example, the experience of certain kinds of social violence has been linked to changes in the epigenetic regulation of genes acting in the hypothalamic-pituitary-adrenal (HPA) axis, which controls reactions to stress. Investigations into the methylation status of the glucocorticoid receptor, which interacts with cortisol and other glucocorticoid hormones to control stress responses, have shown changes in intergenerational epigenetic patterns stemming from the experience of warfare or intimate partner violence in contemporary populations (Mulligan et al. 2012, Radtke et al 2011). While it would be very difficult to distinguish between competing hypotheses for how these regulatory changes within the HPA axis are triggered, these types of epigenetic expectations could be incorporated into existing theoretical frameworks of social violence within bioarchaeology. Research into the regulation of the HPA axis may reveal epigenetic patterns that correlate with forms of gendered violence, regional warfare, or periods of turbulent social change in the ancient world, which could add significantly to

our understanding of individual and population level effects of violence and social transformation (Perez et al, in press; Martin et al, in press).

Epigenetic gene inactivation has also been linked to changes in nutritional status during fetal development. Changes in methylation patterns associated with genes that influence growth and metabolism, such as the Insulin-like Growth Factor 2 (IGF-2), have been observed in individuals who experienced starvation (Heijmans et al 2008). The analysis of these same patterns in prehistoric populations could help illuminate the nutritional statuses of prehistoric populations and help elucidate the effects of events like climate-induced famines in the ancient world.

The use of these methods in conjunction with traditional population genetic approaches could provide a robust way to answer new questions related to human evolution. However, as noted earlier, problems arising from post-mortem damage to aDNA may complicate the ability to accurately reconstruct the kinds of processes highlighted here. Though more research is needed to evaluate the feasibility of moving from the detection of cytosine methylation to the inference of gene activity in prehistoric populations, the possible applications within anthropological genetics are exciting.

Bibliography

Attwood J et al. 2002. DNA methylation and the regulation of gene transcription. Cell Mol. Life Sci. 59(2):241-57.

Bestor T 1998. The host defense function of genomic methylation patterns. Novartis Found. Symp. 214, 187–195.

Boissinot S et al. 2000. L1 (LINE-1) retrotransposon evolution and amplification in recent human history. Mol. Biol. Evol. 17:915-928.

Bolnick and Bonine, unpublished data.

Bolnick D and Smith D. 2007. Migration and Social Structure among the Hopewell: Evidence from Ancient DNA. American Antiquity 72(4):627-644.

Briggs A et al. 2007. Patterns of damage in genomic DNA sequences from a Neandertal. PNAS 104(37):14616-14621.

Briggs A et al. 2009. Removal of deaminated cytosines and detectionof in vivo methylation in ancient DNA. Nuc. Acid Res. 38(6):1-12.

Calvignac S et al. 2008. Ancient DNA evidence for the loss of a highly divergent brown bear clade during historical times. Mol. Ecol. 17:1962-1970.

Campos P et al. 2010. Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (Ovibos moschatus) population dynamics. PNAS 107(12):5675–5680.

Debruyne R et al. 2008. Out of America: Ancient DNA Evidence for a New World Origin of Late Quaternary Woolly Mammoths. Cur. Bio. 18(17):1320-1326.

Gama-Sosa M et al. 1983. Tissue-specific differences in DNA methylation in various mammals. Biochem. Biophys. Acta. 740(2):212-219.

Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes.J Mol Biol. 1987;196:261–82.

Gilbert M et al. 2005. Assessing ancient DNA studies. Trends in Ecol. Evol. 20(10):541-544.

Green R et al. 2010. A Draft Sequence of the Neandertal Genome. Science 328(5979):710-722.

Grunau et al. 2001. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. Nucleic Acids Research 29(13):1-7.

Heijmans B et al. 2008. Persistent epigenetic differences associated with prenatal exposure to famine in humans. PNAS 105(44): 17046–17049.

Hofreiter M et al. 2001. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in Ancient DNA. Nuc. Acids Res. 29(23)4793-4799.

Hofreiter M et al. 2002. Ancient DNA analyses reveal high mitochondrial DNA sequence diversity and parallel morphological evolution of late pleistocene cave bears. Mol. Biol. Evol. 19(8):1244-1250.

Jaenisch R and Bird A. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nature Genet. 33:245-254.

Kaestle F and Horsburgh K. 2002. Ancient DNA in Anthropology: Methods, Applications,and Ethics. Yearbook Phys. Anth. 45:92-130.

Kazazian H and Moran J. 1998. The impact of L1 retrotransposons on the human genome. Nature Genet. 19(1):19-24.

Llamas B et al. 2012. High-Resolution Analysis of Cytosine Methylation in Ancient DNA. PLoS One 7(1):1-6.

Mata-Míguez and Bolnick, unpublished data.

Mata-Míguez J et al. 2012. The genetic impact of aztec imperialism: Ancient mitochondrial DNA evidence from Xaltocan, Mexico. Am. J. Phys. Anth. 149(4):504-516.

McCallum J et al. 2013. Highly Informative Ancient DNA 'Snippets' for New Zealand Moa. PLoS One 8(1):1-8.

Miller W et al. 2008. Sequencing the nuclear genome of the extinct woolly mammoth. Nature 456:387-392.

Munoz-Lopez M et al. 2012. Analysis of LINE-1 expression in human pluripotent cells. Methods Mol. Biol. 873:113-125.

Myers J et al. 2002. A comprehensive analysis of recently integrated human Ta L1 elements. Am. J. Hum. Genet. 71(2):312-326.

Orlando L et al. 2002. Ancient DNA and the population genetics of cave bears (Ursus spelaeus) through space and time. Mol. Biol. Evol. 19(11)1920-1933.

Paabo S et al. 2004. Genetic Analyses from Ancient DNA. Annu. Rev. Genet. 38:645-679.

Portela A and Esteller M. 2010. Epigenetic modifications and human disease. Nature Biotech. 28(10):1057-1068.

R Core Team. 2013. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.

Reich D et al. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468:1053-1060.

Reik W. 2007. Stability and flexibility of epigenetic gene regulation in mammalian development. Nature 447(7143):425-432.

Sheen F et al. 2000. Reading between the LINEs: Human genomic variation induced by LINE-1 retrotransposition. Genome Res. 10(10):1496-1508.

Stiller M et al. 2010. Withering Away: 25,000 Years of Genetic Decline Preceded Cave Bear Extinction. Mol. Biol. Evol. 27(5):975-978.

Sved J, Bird A. The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. Proc Natl Acad Sci U S A. 1990;87:4692–6.

Takai D, Jones PA. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci U S A. 2002;99:3740–5.

Villanea et al., in press.

Willerslev E and Cooper A. 2005. Ancient DNA. Proc. Royal Soc. Biol. Sci. 272(1558):3-16.

Yoder J et al. 1997. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. 13(8)335-340.

Zhang F et al. 2011. Dietary Patterns Are Associated with Levels of Global Genomic DNA Methylation in a Cancer-Free Population. J. Nutrition 141(6): 1165-1171.